Pertinence des pseudo-mots dans l'évaluation de l'intelligibilité : Effet du nombre ou du caractère non lexical ?

Marie Rebourg, Muriel Lalain, Alain Ghio, Corinne Fredouille, Nicolas Fakhry, Virginie Woisard

Pertinence des pseudo-mots dans l'évaluation de l'intelligibilité : Effet du nombre ou du caractère non lexical ?

Marie Rebourg[1], Muriel Lalain[1], Alain Ghio[1], Corinne Fredouille[2], Nicolas Fakhry[1, 3], Virginie Woisard[4]

(1) Aix-Marseille Univ, CNRS, LPL, UMR 7309, Aix-en-Provence, France
(2) Avignon Université, Laboratoire Informatique d'Avignon, France
(3) Service ORL, APHM, La Conception, Marseille, France
(4) Service ORL, CHU Larrey, UT2J Laboratoire de NeuroPsychoLinguistique, Toulouse, France

marie.rebourg@univ-amu.fr, muriel.lalain@univ-amu.fr, alain.ghio@univ-amu.fr, corinne.fredouille@univ-avignon.fr, nicolas.fakhry@ap-hm.fr, woisard.v@chu-toulouse.fr

RESUME

La perte d'intelligibilité constitue une plainte récurrente des patients traités pour un cancer de la cavité buccale ou de l'oropharynx. La notion d'intelligibilité par son aspect factorielle est complexe à définir, mais aussi, par extension, à évaluer avec précision. Les différents matériaux utilisés dans ces évaluations sont connus pour montrer des effets d'apprentissages imputables aux listes d'items courtes et fermées, présentes dans les batteries de tests classiques. Dans cette étude, nous évaluons l'effet d'apprentissage du matériel linguistique en comparant l'évolution des scores d'intelligibilité calculés à partir de la transcription de mots et de pseudo-mots, présentés en proportion équivalente, soit la répétition de 50 mots vs de 52 pseudo-mots. Nos résultats montrent un effet d'apprentissage des pseudo-mots lorsqu'ils sont répétés, dans les mêmes proportions que celui observé sur les mots. Ainsi, c'est la quantité de pseudo-mots qui permet de neutraliser l'effet d'apprentissage du matériel linguistique dans une évaluation de l'intelligibilité.

ABSTRACT

Relevance of pseudowords in the assessment of intelligibility: Effect of number or non-lexical character ?

Loss of intelligibility is a recurring complaint among patients treated for cancer of the oral cavity or oropharynx. The notion of intelligibility due to its multifactorial aspect is complex to define, but also, by extension, to evaluate with precision. The different materials used in these evaluations are known to show learning effects attributable to the short and closed lists of items, present in traditional test batteries. In this study, we evaluate the learning effect of linguistic material by comparing the evolution of intelligibility scores calculated from the transcription of words and pseudo-words, presented in equivalent proportion, i.e. the repetition of 50 words vs. 52 pseudowords. Our results show a learning effect on pseudowords when they are repeated, in the same proportions as that observed on words. Thus, it is the quantity of pseudo-words which makes it possible to neutralize the learning effect of linguistic material in an evaluation of intelligibility.

MOTS-CLÉS : Phonétique Clinique, Intelligibilité, Trouble de la Production de la Parole, Cancer VADS

KEYWORDS : Clinical phonetic, Intelligibility, Speech disorders, Head and Neck cancer

1 L'intelligibilité

Les traitements dont bénéficient les patients dans le cadre de la prise en charge des cancers de la cavité buccale et de l'oropharynx sont connus pour porter atteinte aux fonctionnements anatomiques de l'appareil phonatoire. Il en résulte une perte d'intelligibilité, qui constitue une plainte récurrente chez ces patients, en perte d'autonomie communicationnelle. L'évaluation de cette composante linguistique multidimensionnelle est essentielle dans le parcours de soin du patient, puisqu'elle permet de mesurer le handicap à la communication en évaluant les composantes dégradées/préservées et de mesurer l'effet du traitement préalablement établi.

L'intelligibilité, par sa dimension multiple et factorielle, représente une notion complexe à définir et à circonscrire. Plusieurs définitions et approches ont été proposées, telles que « le degré de précision avec lequel un message est compris par un auditeur » (Yorkston, Dowden et Beukelman, 1992). Toutefois, ces définitions intègrent la notion de compréhension, entrainant une confusion entre les différents niveaux d'informations linguistiques qui doivent être ciblés par la notion d'intelligibilité. Le concept de compréhension réfère à l'intégration de l'ensemble des informations pertinentes indépendantes et dépendantes du signal permettant de comprendre, de saisir le sens d'un énoncé oral en situation de communication (Lindblom, 1990). Cette acception soutient ainsi l'idée selon laquelle les processus de bas niveaux supportent des informations « dépendantes du signal » alors que les informations « indépendantes du signal » sont liées aux processus de haut niveau (Hustad, Jones et Dailey, 2003). Ainsi, la définition de l'intelligibilité doit être circonscrite au bas niveau, comme proposé par Hustad (2008) selon laquelle l'intelligibilité fait référence à « la façon dont le signal acoustique d'un locuteur peut être récupéré avec précision par un auditeur ». Cette approche permet d'évacuer la dimension de compréhensibilité (haut + bas niveaux) tout en recentrant le concept d'intelligibilité autour des informations acoustico-phonétiques, dépendantes du signal acoustique (bas niveau).

L'évaluation, la mesure, de l'intelligibilité doit donc être rigoureusement contrôlée de façon à déterminer quel matériel linguistique permet d'évaluer l'intelligibilité de quel niveau linguistique. En d'autres termes, la sélection du matériel linguistique utilisé pour évaluer l'intelligibilité doit être conditionnée par la définition précise du niveau linguistique visé par la mesure.

1.1 Mesurer l'intelligibilité – études et contexte clinique

Les études récentes concernant la mesure d'intelligibilité portent tant sur le matériel linguistique que sur la méthode d'évaluation, mais aussi sur la méthode de calcul choisie. Elles s'accordent sur la nécessité de mesurer précisément quelle composante linguistique est évaluée en fonction des différents matériaux linguistiques. Dans la lignée de Ganzeb*oom et* al. (2016), *Xue et* al. (2021) suggèrent que différentes constructions de l'intelligibilité sont nécessairement reflétées par les mesures d'intelligibilité selon qu'elles emploient des échelles analogiques ou des évaluations par transcriptions, mais aussi selon le matériel linguistique employé dans cette évaluation, distinguant les phrases, les mots et les pseudo-mots. Cette proposition ancre par conséquent l'aspect multi-dimensionnel de l'intelligibilité.

Ainsi, les études considérant des ensembles de phrases, questionnant la pertinence de ce matériel linguistique pour évaluer l'intelligibilité, ont montré que l'intégration du contexte menait davantage à une évaluation de la compréhension, puisque celles-ci intègrent les informations de haut et bas niveaux (Yorkston, Strand et Kennedy, 1996; Ganzeb*oom et* al., 2016; Xue, R. Ho*ut, et* al., 2021). Cela représente donc une évaluation perceptive du déficit global. De plus, en comparaison avec les phrases, les listes de mots se montrent comme de meilleurs candidats (*Xue et* al., 2023). Néanmoins, les listes de mots disponibles dans les batteries d'évaluation clinique

(BECD (Auzou et Rolland-Monnoury, 2006), FDA2 (Bl*anc et* al., 2014)) se montrent facilement mémorisables de par leur caractère court et fermé. La répétition de la tâche par le clinicien et un nombre d'items restreints favorisent les effets d'apprentissages et de restauration lexicale. Ceux-ci sont donc inhérents aux matériaux linguistiques de nature lexicale. De plus, ces listes se montrent peu contrôlables au regard de l'occurrence et de la position des phonèmes au sein des items. Restent donc les pseudo-mots, items non lexicaux, n'étant pas porteurs de sens et suivant les règles phonotactiques de la langue, qui se positionnent comme item candidat idéal et pertinent. Ils permettent de s'affranchir de l'intégration des informations de haut niveau, au profit des unités de bas niveau, forçant ainsi le décodage acoustico-phonétique des sons de parole. Ils peuvent être générés de façon automatique, selon un ensemble de contraintes administrant l'occurrence et la position des phonèmes. Ainsi, le seul effet de restauration qui pourrait être attendu avec ce type de matériel linguistique concernerait le niveau du phonème, soit une reconstruction basée sur les connaissances phonologiques d'organisation des unités phonémiques de la langue par l'auditeur.

Concernant les différentes méthodes d'attribution et de calcul des scores, elles peuvent reposer sur des échelles de mesure graduée (Lickert), des échelles analogiques visuelles, ou des transcriptions. Les études les plus récentes montrent que les scores les plus fiables sont obtenus par transcriptions orthographiques (*Xue et* al., 2023). Les différentes échelles prennent en compte des paramètres subjectifs, qui, mêmes s'ils sont précisés, restent empreints d'une grande variabilité, dépendant de l'auditeur-évaluateur. De plus, ces scores perceptifs globaux se montrent peu pertinents dans l'identification des composantes dégradées/préservées et des niveaux linguistiques touchés par ces altérations. Ainsi, les scores calculés de façon automatique, selon une méthodologie précise, sont perçus comme plus fiables et plus objectifs.

En ce sens, une tâche de Décodage acoustico-phonétique (ci-après DAP) permettant de collecter des transcriptions orthographiques, basée sur la perception de pseudo-mots produits, doublée d'une méthode de calcul innovante, basée sur la théorie des traits distinctifs, a été développée dans le cadre du projet de recherche C2SI (Carcinologic Speech Severity Index, Institut National pour le Cancer n°2014-135) (Astésano *et al.*, 2018; Lalain *et al.*, 2020; Woisard *et al.*, 2021). Celle-ci permet le calcul d'un score analytique en termes de distance à la cible, en nombre de traits moyens altérés par phonème (Ghio, Lalain, Giusti, *et al.*, 2020). La mise à l'épreuve de cette tâche a montré sa pertinence pour distinguer deux groupes de locuteurs – patient vs contrôle (Ghio *et al.*, 2018). Elle a également montré que le recours aux pseudo-mots permet d'attribuer des scores stables au cours du temps, contrairement à l'utilisation de mots ; ceci suggérant que les effets d'apprentissage du matériel linguistique, rencontrés dans les batteries de tests classiques, peuvent être neutralisés par une grande diversité au sein des listes de stimuli (Rebo*urg et* al., 2019; Lal*ain et* al., 2022). De plus, l'évaluation de la pertinence de cette tâche a également montré que l'effet d'expertise auditive des cliniciens était préservé. En moyenne, les scores calculés à partir des transcriptions de ces auditeurs sont plus bas que ceux des auditeurs naïfs, suggérant qu'ils sont de meilleurs décodeurs (Rebo*urg et* al., 2020). La méthode de calcul employée a également montré sa pertinence pour proposer une mesure fine et fiable représentative du déficit articulatoire et par extension, du handicap communicationnel relatif à la qualité de vie du patient.

Enfin, les études les plus récentes, au regard des différentes mesures proposées pour évaluer l'intelligibilité en contexte clinique, montrent que les mesures portant au niveau du phonème sont pertinentes pour détecter les erreurs articulatoires, dans le cadre des dysarthries (*Xue et* al., 2023). Et donc, par extension, probablement pour apprécier finement les séquelles articulatoires, et leur évolution, dans le cadre de la prise en charge clinique et orthophonique de patients après un cancer de la cavité buccale ou de l'oropharynx.

Dans la présente étude, nous interrogeons les résultats précédemment obtenus (Rebo*urg et* al., 2020) qui montrent que l'utilisation d'un très grand nombre de pseudo-mots permet de neutraliser

les effets d'apprentissage du matériel linguistique couramment utilisé pour évaluer l'intelligibilité. La présente étude questionne l'origine de cet effet : est-il davantage lié au nombre d'items, en termes de quantité, ou aux pseudo-mots en tant qu'unité linguistique non lexicale, soit sa qualité ?

2 Méthodologie

Afin de satisfaire à ces différentes questions de recherche nous avons constitué deux corpus, basés sur la production de mots (M) et de pseudo-mots (PM) isolés. Ces corpus ont été utilisés dans des tests de jugement perceptif de l'intelligibilité, visant à obtenir les transcriptions orthographiques des stimuli perçus. Celles-ci, après phonétisation, permettent le calcul des scores de Déviation Phonologique Perçue (DPP) reflétant le nombre de traits (Jakobson, Fant et Halle, 1952) moyens altérés par phonème (Ghio, Lalain, Giusti, *et al.*, 2020)

2.1 Corpus

Pour mener à bien ces travaux de recherche, deux corpus ont été constitués. (i) Un premier corpus de production de mots isolés. Il comprend les enregistrements de 20 locuteurs : 10 patients traités pour un cancer de la cavité buccale ou de l'oropharynx (Toulouse (Balaguer, 2021)) et 10 sujets contrôles (Aix-en-Provence (Rebourg, 2022)), appariés en âge et en sexe. Ils ont été enregistrés lors de la production de la liste de 50 mots de la BECD (Auzou et Rolland-Monnoury, 2006) couramment utilisée pour évaluer l'intelligibilité en contexte clinique. (ii) Un second corpus, comprenant des productions de pseudo-mots isolés, a été constitué auprès d'un autre groupe de 20 locuteurs. Également, 10 patients traités pour un cancer de la cavité buccale ou de l'oropharynx (Toulouse (Balaguer, 2021)), 10 sujets contrôles (Aix-en-Provence (Rebourg, 2022)), appareillés en âge et en sexe. Ils ont tous été enregistrés pendant la production d'une **seule et unique liste de 52 pseudo-mots**, extraite du matériel linguistique développé dans le cadre du projet C2SI, spécifiquement élaboré pour une tâche perceptive de décodage acoustico-phonétique. Ces enregistrements, traités et préparés pour satisfaire les conditions expérimentales, comptabilisent 1000 stimuli dans le corpus de Mots et 1040 stimuli pour le corpus de Pseudo-mots.

Ces travaux de recherche emploient un protocole expérimental déjà éprouvé (Ghio *et al.*, 2018; Rebourg, 2018, 2022; Rebourg *et al.*, 2019, 2020; Ghio, Lalain, Rebourg, *et al.*, 2020; Lalain *et al.*, 2022). Afin de limiter le nombre de stimuli présentés et transcrits par chaque auditeur, dans le test de perception, les groupes de locuteurs constitutifs des corpus sus-présentés ont été divisés en deux sous-groupes (A et B). Chacun de ces sous-groupes comprend 5 locuteurs patients et 5 contrôles. Ainsi, chacun des corpus (Mots et Pseudo-mots) a été divisé en 2 sous-corpus selon les groupes de locuteurs A et B. Ils comprennent respectivement 500 stimuli Mots liste BECD) et 520 stimuli Pseudo-mots (tâche de DAP, projet C2SI).
Les 2 x 500 stimuli du corpus de mots et les 2 x 520 stimuli du corpus de pseudo-mots ont respectivement été divisés en 2 x 3 listes. Chaque liste de mots (BECD) est constituée de 167 productions et chaque liste de pseudo-mots (DAP) est constituée 174 productions. Soit un total de 6 listes pour chaque corpus et 3 listes pour chaque sous-corpus (groupes de locuteurs A et B). Les listes 1 à 3 de chacun des corpus (Mots et Pseudo-mots) sont produites par le groupe de locuteurs A et les listes 4 à 6 par le groupe de locuteurs B.

2.2 Design expérimental

Dans ce test de perception, les auditeurs écoutent les stimuli produits dans un casque et transcrivent orthographiquement sur un clavier d'ordinateur ce qu'ils entendent. Chaque auditeur écoute et transcrit 3 listes BECD et 3 listes DAP, produites par un groupe de locuteurs (A ou B). Chaque liste est transcrite par 3 auditeurs différents. L'ordre de présentation de chaque liste a été contrôlé de sorte que chaque liste occupe chaque position possible. Nous obtenons donc 3

transcriptions de 3 auditeurs différents dans chacun des 3 ordres de présentation, soit 9 transcriptions par liste.

			T1	T2	T3
Groupes de locuteurs A	3 auditeurs ≠	Mots BECD	1	2	3
		Pseudo-mots	1	2	3
	3 auditeurs ≠	Mots BECD	2	3	1
		Pseudo-mots	2	3	1
	3 auditeurs ≠	Mots BECD	3	1	2
		Pseudo-mots	3	1	2

TABLE 1 : Extrait récapitulatif du design expérimental du test de jugement perceptif par DAP, chaque chiffre des colonnes T1 à T3 correspond aux identifiants des différentes listes.

Ce protocole expérimental a permis de mesurer l'évolution des scores PPD au cours de la répétition de 50 mots et de 52 pseudo-mots; scores calculés à partir des transcriptions des auditeurs dans un test de jugement perceptif de l'intelligibilité par décodage acoustico-phonétique (DAP).

2.3 Test de perception

A travers la base de données de volontaires participant aux expériences scientifiques du Centre d'Expérimentation sur la Parole (CEP) du LPL, nous avons recruté 18 auditeurs naïfs. Tous natifs de langue française, sans problème de vue ou d'audition non corrigés et ayant un bon niveau en orthographe. Chaque participant a débuté la tâche par la transcription des productions de mots isolés, puis l'a poursuivie avec la transcription des productions de pseudo-mots. Un auditeur évalue donc 10 locuteurs (5 patients et 5 contrôles), soit un total de 1020 stimuli (500 mots et 520 pseudo-mots). Les auditeurs ont été dédommagés en ticket Kadeos.
Ils ont reçu la consigne de toujours proposer une transcription, qui soit au plus près de ce qu'ils ont perçu et identifié, en respectant les règles orthographiques du français. Ce test de jugement perceptif de l'intelligibilité a été conduit au CEP (http://cep.lpl-aix.fr/), à l'aide de la station de perception PercEval (An*dré et* al., 2003). Ce design expérimental a été élaboré pour évaluer l'évolution des scores de Déviation Phonologique Perçue au cours du temps de la tâche en fonction du matériel linguistique utilisé (Mots – Pseudo-mots) à parts égales.

2.4 Traitement des données

Afin de pouvoir analyser statistiquement ces données, collectées lors des tests de jugement perceptif de l'intelligibilité par DAP, plusieurs traitements sont nécessaires. Les données brutes, telles que présentées par le logiciel d'expérimentation PercEval (An*dré et* al., 2003), sont des transcriptions orthographiques. Différentes opérations de pré-traitement sont effectuées afin que le format des données soit compatible avec les outils de calcul des scores DPP. Ces transcriptions orthographiques sont phonétisées en deux étapes (LIA-Phon, (Béchet, 2001) et Lexique.org) et sont ainsi compatibles avec la matrice de confusion utilisée pour calculer les scores de Déviation Phonologique Perçue (DPP).
Cette matrice de confusion repose sur l'attribution d'un coût, en termes de distance, basé sur la théorie des traits distinctifs (Jakobson, Fant et Halle, 1952). Elle a été développée par A. Ghio dans le cadre de l'analyse des données DAP (Ghio *et al.*, 2018; Ghio, Lalain, Giusti, *et al.*, 2020) du projet de recherche C2SI (Astésano *et al.*, 2018; Ghio, Lalain, Giusti, *et al.*, 2020). Elle permet le calcul de distance d'édition entre deux chaînes de caractères phonétiques par l'algorithme de Wagner Fischer. Celui-ci intègre la distance de Levenshtein, qui considère trois opérations

d'éditions élémentaires : la suppression, l'insertion ou la substitution d'un caractère. Ceci permet de considérer les altérations sur les deux axes syntagmatique et paradigmatique. En d'autres termes, cette méthode permet d'attribuer un score de distance entre les transcriptions phonétiques des cibles qui devaient être prononcées par les locuteurs et les transcriptions effectives phonétisées des auditeurs.
Le calcul des scores de distances cumulées repose sur la division du score donné par la matrice de confusion par le nombre de caractères de la cible phonétique. Nous obtenons alors des scores de Déviation Phonologique Perçue (DPP) en termes de distance cumulée à la cible, qui représentent le nombre moyen de traits altérés par phonème. De plus, pour chaque auditeur, un numéro a été assigné à chaque item en fonction de l'ordre de la passation, de 1 à 500 pour les mots et de 1 à 520 pour les pseudo-mots. Nous comparons donc l'évolution des scores au cours de la tâche entre deux matériaux linguistiques répétés à occurrences égales, 50 mots et 52 pseudo-mots.

3 Résultats

Cette étude a été menée afin d'évaluer si la neutralisation des effets d'apprentissage (Rebo*urg et* al., 2020) tenait davantage à la qualité du matériel linguistique : Mots vs Pseudo-mots, ou à la quantité de pseudo-mots différents. Pour rappel, dans l'expérience précédente une même liste de 50 mots et 2 listes de 52 pseudo-mots, soit 50 mots répétés 20 fois et 2080 pseudo-mots différents, constituaient les stimuli de l'expérience. Il s'agit ici de confronter la répétition d'une liste de 50 mots à la répétition d'une liste de 52 pseudo-mots.
Une analyse de variance (ANOVA) a été conduite, afin de tester les effets simples entre l'évolution des scores PPD moyen (VD) au cours de la tâche et les différents matériaux linguistiques (Mots vs Pseudo-mots). Comme précédemment (Rebo*urg et* al., 2020; Rebourg, 2022), nos résultats concernant le matériel linguistique lexical (Mots) confirment un effet de l'ordre des items dans la passation ($F(1,498) = 38.72$, $r = 0.004$, $p < 0.001$). Autrement dit, le score moyen baisse significativement au cours de la répétition de la tâche montrant un effet d'apprentissage du matériel linguistique lexical. Les résultats obtenus pour les pseudo-mots montrent également un effet significatif de l'évolution à la baisse du score DPP moyen au cours de la répétition de la tâche ($F(1,518) = 64.78$, $r = 0.006$, $p < 0.001$). Cela suggère qu'un effet d'apprentissage est induit par la répétition d'un même matériau linguistique, quelle que soit sa nature, lexicale ou non lexicale.

Ces résultats sont illustrés par la Figure 1 qui montre une relation affine forte et négative entre les scores moyens d'intelligibilité et la position de l'item dans la passation, pour les Mots et les Pseudo-mots. Cette figure se lit comme suit : en ordonnée les scores de Déviation Phonologique Perçue (PPD) ; en abscisse le numéro attribué à l'item en fonction de son ordre dans la passation (1 à 500 pour les Mots, 1 à 520 pour les Pseudo-mots) ; chaque point représente la moyenne des scores moyens attribués aux items en fonction de leur ordre dans la passation ; les droites de régression linéaires montrent l'évolution globale des scores au cours de la tâche.

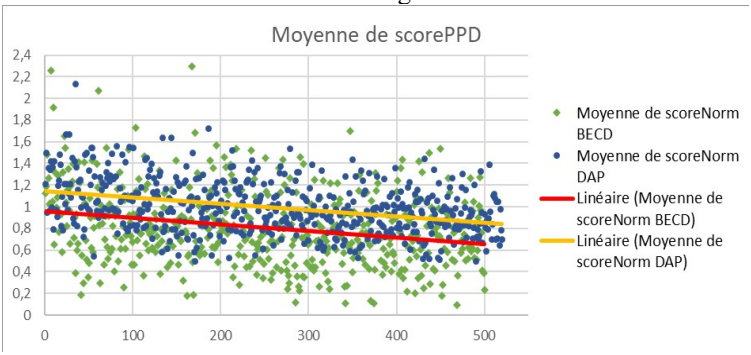

FIGURE 1 : Représentation en nuage de points des scores d'intelligibilité moyen en fonction de la position de l'item dans la passation, du type de matériel linguistique, Mots et Pseudo-Mots

La baisse de scores DPP moyens au cours de la tâche se traduit par un score moyen plus faible à la fin de la tâche, soit au 500/520ème mots/pseudo-mots, par rapport aux premiers stimuli transcrits. Pour les mots, losanges verts, droite de régression linéaire rouge, les scores moyens passent de 0,97 pour la première transcription à 0,64 pour la 500ème transcription, soit un abaissement de 0,33 traits par phonème. Cela représente une baisse de 34% du score moyen pour 500 essais. Pour les pseudo-mots, points bleus, droite de régression linéaire jaune, les scores moyens passent de 1,19 pour la première transcription à 0,86 pour la 520ème transcription, soit un abaissement de 0,33 traits par phonème, égale à celui des mots. Cela représente une baisse de 28% du score moyen. Cet effet de baisse significative des scores moyens représente un effet d'amélioration de 0,00066 trait par phonème par essai, pour les deux matériaux linguistiques. Cela révèle un effet d'apprentissage du matériel linguistique au cours de la tâche.

De plus, les scores moyens PPD calculés pour les Pseudo-mots ($\mu = 0.99$) sont significativement plus élevés ($p < 0.001$) que ceux obtenus avec une évaluation avec les Mots ($\mu = 0.80$). Cela suggère qu'une évaluation basée sur des pseudo-mots est plus stricte qu'une évaluation basée sur des mots de lexique. La taille de cet effet mesurée avec un coefficient d de Cohen ($d = -0.153$) nous indique que cette différence est faible, représentant 0.2 trait moyen d'écart par phonème.

4 Conclusion - discussion

Ces résultats révèlent un effet d'apprentissage du matériel linguistique lors de la répétition d'une tâche comprenant des listes d'items courtes et fermées, quelle que soit la nature du matériel linguistique lexical et non lexical. Mis en perspectives avec les résultats de notre précédente expérience, ces présents résultats montrent que c'est la quantité de pseudo-mots, qui permet de neutraliser cet effet d'apprentissage. En effet, aucun effet d'apprentissage au cours de la tâche n'avait été révélé par la présentation de 2080 pseudo-mots différents, contrairement à celui mis en avant ici, lors de la répétition d'une liste de 52 pseudo-mots. Ces résultats suggèrent que dans le cadre de l'évaluation clinique de la parole, un grand répertoire de pseudo-mots constitue un matériel linguistique pertinent pour neutraliser les effets d'apprentissage des listes d'items, biais fréquent des batteries de tests classiques d'évaluation de l'intelligibilité. De plus, ces résultats contribuent à valider les critères d'objectivité et de pertinence de la tâche DAP (C2SI).

Nos résultats montrent une différence significative de scores moyens entre les deux matériaux linguistiques, plus élevée pour les Pseudo-mots que pour les Mots, de 0.2 trait moyen d'écart par phonème. Cela suggère que la nouveauté de ces formes linguistiques non lexicales, jamais perçues auparavant, contrecarre l'effet de restauration lexicale. Les scores, en moyenne moins élevés, obtenus avec des mots peuvent s'expliquer par le recours aux informations de haut niveau. L'auditeur fait appel à son lexique mental dans lequel il a stocké ces mots de lexique, probablement déjà rencontrés au cours de son expérience linguistique, facilitant ainsi un effet de restauration lexicale, susceptible d'être délétère dans le cadre de l'évaluation clinique de la parole.

Néanmoins, nos résultats montrent que les scores moyens baissent au cours de la tâche pour les deux matériaux linguistiques. Et ce dans une amplitude exactement identique, 0,33 traits entre le premier et le dernier stimuli transcrit, soit en moyenne 0,00066 trait par phonème, par stimuli transcrit. Cette baisse des scores moyens proportionnellement identique entre les Mots et les Pseudo-mots suggère que les mêmes mécanismes se mettent en place lors de la perception d'unités linguistiques non lexicales telles que les pseudo-mots, analogues à ceux de la perception de mots

du lexique. L'auditeur tient à jour ses connaissances linguistiques et se montre capable d'intégrer, dans son stock de connaissances lexicales, des unités phoniques auxquelles aucun sens n'est attribué. Il réinvestit ses connaissances de haut niveau en ajoutant des stimuli à son répertoire de formes linguistiques perçues. Ceci expliquerait l'effet d'apprentissage révélé dans notre expérience. De plus, cet effet d'apprentissage des pseudo-mots questionne l'emploi d'un très grand répertoire de mots pour une évaluation perceptive. En effet, Si seul un grand nombre d'items évite les biais perceptifs, il peut être tentant de choisir des mots qui sont des éléments linguistiques familiers contrairement aux pseudo-mots considérés comme plus artificiels. Ainsi, on pourrait envisager un recours à tous les mots CVCV (4087), CCVCV (977), CVCCV (1648), CCVCCV (150) présents dans la base lexique.org. Cela représente 6862 mots, ce qui pourrait être une taille suffisante. Cependant, avec un tel dictionnaire, il serait très difficile d'obtenir des listes équivalentes de 50 mots en termes de contenu et d'équilibre phonétique, et ne permettraient pas de contrôler les effets de restauration lexicale. D'autre part, il serait aussi compliqué de maitriser des contraintes de fréquence d'apparition (mots rares vs fréquents), alors que cet aspect est crucial dans la perception du lexique (Vitevitch et Luce, 1999). Seul le recours aux pseudo mots permet l'obtention de listes phonétiquement équivalentes (contrôle de l'occurrence et de la position des phonèmes) et où la contrainte de la fréquence d'occurrence est annihilée par la nature inédite de ces unités.

De plus, la qualité des pseudo-mots, en tant qu'items linguistiques non lexicaux, permet de contrôler les effets de restauration lexicale en orientant l'auditeur dans le sens de la restauration phonémique. Les connaissances implicites de l'auditeur à propos de règles phonologiques de sa langue, en tant que système régi par des ensembles de règles d'ordonnancement et d'organisation des sons, lui confèrent cette capacité. De plus les pseudo-mots, en tant qu'unités linguistiques non lexicales, présentent l'avantage de pouvoir être générés en très grande quantité, neutralisant les effets d'apprentissage et de ne jamais avoir été entendus auparavant, neutralisant les effets de restauration lexicale, en centrant la tâche de l'auditeur sur le décodage des sons de parole, soit du décodage acoustico-phonétique. Ces qualités suggèrent qu'ils constituent des unités pertinentes pour l'évaluation des composantes phonétiques articulatoires et acoustiques préservées/dégradées, dans le cadre clinique.

L'ensemble de ces résultats souligne également la complexité de l'ensemble des mécanismes de perception de la parole. Ainsi, la compréhension et l'appréhension convenable des liens entre les concepts linguistiques fondamentaux et les méthodes d'évaluation les plus appropriées pour cibler ces niveaux constitue l'un des apports à la linguistique fondamentale. Celui-ci est permis par l'observation à travers le prisme des déficits et altérations de la parole, tout en soutenant l'objectif de l'évaluation en contexte clinique.

Reste à établir et déterminer avec précision les degrés de corrélation entre les différentes mesures, tant dans les méthodes (tâche de l'auditeur : notation par échelles, par transcriptions), que dans les procédures de calcul des scores (globale vs analytique) ainsi qu'entre les différents matériaux linguistiques (énoncés, phrases, mots, pseudo-mots). En ce sens, le développement d'un index d'évaluation de la parole qui présenterait synthétiquement les différents matériaux linguistiques associés avec les niveaux linguistiques évalués, en fonction des différentes méthodes de scorage, constituerait un outil essentiel tant à la pratique clinique qu'à la linguistique fondamentale.

Références

ANDRÉ, C. *ET AL.* (2003): «PERCEVAL: a Computer-Driven System for Experimentation on Auditory and Visual Perception», in *XVth ICPhS*. *ICPhS*, Barcelone, Espagne, p. 1421-1424.

ASTÉSANO, C. *ET AL.* (2018): «Carcinologic Speech Severity Index Project: A Database of Speech Disorder Productions to Assess Quality of Life Related to Speech After Cancer», in. *Language Resources and Evaluation Conference*, Miyazaki, p. 7.

AUZOU, P. ET ROLLAND-MONNOURY, V. (2006): *Batterie d'évaluation clinique de la dysarthrie*. Ortho Edition. France: ORTHO.

BALAGUER, M. (2021): *Mesure de l'altération de la communication par analyses automatiques de la parole spontanée après traitement d'un cancer oral ou oropharyngé*.

BECHET, F. (2001): «LIA PHON : Un système complet de phonétisation de textes», *Traitement Automatique des Langues, TAL*. (TAL - ATALA), 42(1), p. 47-67.

BLANC, E. *ET AL.* (2014): «Adaptation en français du test d'intelligibilité de la version révisée du « Frenchay Dysarthria Assessment » (FDA-2)», in *Congrès de la Société Française de Phoniatrie*. Paris, France. Disponible sur: https://hal.archives-ouvertes.fr/hal-01615204.

GANZEBOOM, M. *ET AL.* (2016): «Intelligibility of Disordered Speech: Global and Detailed Scores», in, p. 2503-2507. doi: 10.21437/Interspeech.2016-1448.

GHIO, A. *ET AL.* (2018): «Une mesure d'intelligibilité par décodage acoustico-phonétique de pseudo-mots dans le cas de parole atypique», in *XXXIIe Journées d'Etudes sur la Parole*. Aix-en-Provence, France: ISCA, p. 285-293. doi: 10.21437/jep.2018-33.

GHIO, A., LALAIN, M., GIUSTI, L., *ET AL.* (2020): «How to Compare Automatically Two Phonological Strings: Application to Intelligibility Measurement in the Case of Atypical Speech», in *12th Conference on Language Resources and Evaluation (LREC 2020)*. Marseille, France: ELRA, p. 1682-1687. Disponible sur: https://hal.archives-ouvertes.fr/hal-02482615.

GHIO, A., LALAIN, M., REBOURG, M., *ET AL.* (2020): «Testing intelligibility through acoustic-phonetic decoding of pseudowords : Construct and concurrent validation based on patients with head and neck cancers», *Journal of Communication Disorders*.

HUSTAD, K. C. (2008): «The Relationship Between Listener Comprehension and Intelligibility Scores for Speakers With Dysarthria», *Journal of Speech, Language, and Hearing Research*, 51(3), p. 562-573. doi: 10.1044/1092-4388(2008/040).

HUSTAD, K. C., JONES, T. ET DAILEY, S. (2003): «Implementing Speech Supplementation Strategies: Effects on Intelligibility and Speech Rate of Individuals With Chronic Severe Dysarthria», *Journal of Speech, Language, and Hearing Research*, 46(2), p. 462-474. doi: 10.1044/1092-4388(2003/038).

JAKOBSON, R., FANT, C. G. M. ET HALLE, M. (1952): *Preliminaries to speech analysis: the distinctive features and their correlates*. Cambridge, Etats-Unis d'Amérique: Acoustics Laboratory, Massachusetts Institute of Technology.

LALAIN, M. *ET AL.* (2020): «Design and Development of a Speech Intelligibility Test Based on Pseudowords in French: Why and How?», *Journal of Speech, Language, and Hearing Research*, 63(7), p. 2070-2083. doi: 10.1044/2020_JSLHR-19-00088.

LALAIN, M. *ET AL.* (2022): «Prédiction du degré d'altération de l'intelligibilité chez des patients traités pour un cancer de la cavité buccale ou de l'oropharynx», in *32ème édition des Journées d'Etudes sur la Parole*. Noirmoutier, France.

LINDBLOM, B. (1990): «On the communication process: Speaker listener interaction and the development of speech.», in *Augmentative and Alternative Communication*. (6), p. 220-230.

REBOURG, M. (2018): *Validation d'une tâche de Décodage Acoustico Phonétique : Lexicalisation, mémorisation, familiarisation*. Mémoire de Master 2 - Sciences du langage - Linguistique expérimentale. Aix-Marseille Université.

REBOURG, M. *ET AL.* (2019): «Pertinence de l'utilisation de non mots pour évaluer l'intelligibilité», in *Journées de Phonétique Clinique*. Mons, Belgium (Questions de Phonétique Clinique), p. 172. Disponible sur: https://hal.archives-ouvertes.fr/hal-02098845.

REBOURG, M. *ET AL.* (2020): «Évaluer l'intelligibilité, mots ou pseudo-mots ? Comparaison entre deux groupes d'auditeurs», in *6e conférence conjointe Journées d'Études sur la Parole (JEP, 31e*

édition), (TALN, 27e édition), (RÉCITAL, 22e édition). Nancy, France: ATALA, p. 543-551. Disponible sur: https://hal.archives-ouvertes.fr/hal-02798584.
REBOURG, M. (2022): *Évaluation de l'intelligibilité après un cancer ORL : approche perceptive par décodage acoustico-phonétique et mesures acoustiques*. These de doctorat. Aix-Marseille. Disponible sur: https://www.theses.fr/2022AIXM0247 (Consulté le: 10 février 2024).
VITEVITCH, M. S. ET LUCE, P. A. (1999): «Probabilistic Phonotactics and Neighborhood Activation in Spoken Word Recognition», *Journal of Memory and Language*, 40(3), p. 374-408. doi: 10.1006/jmla.1998.2618.
WOISARD, V. *ET AL.* (2021): «C2SI corpus: a database of speech disorder productions to assess intelligibility and quality of life in head and neck cancers», *Language Resources and Evaluation*. Springer Verlag, 55(1), p. 173-190. doi: 10.1007/s10579-020-09496-3.
XUE, W., HOUT, R., *ET AL.* (2021): «Assessing speech intelligibility of pathological speech: test types, ratings and transcription measures», *Clinical Linguistics & Phonetics*, 37. doi: 10.1080/02699206.2021.2009918.
XUE, W., HOUT, R. V., *ET AL.* (2021): «Speech Intelligibility of Dysarthric Speech: Human Scores and Acoustic-Phonetic Features», in *Interspeech 2021*. *Interspeech 2021*, ISCA, p. 2911-2915. doi: 10.21437/Interspeech.2021-1189.
XUE, W. *ET AL.* (2023): «Assessing speech intelligibility of pathological speech in sentences and word lists: The contribution of phoneme-level measures», *Journal of Communication Disorders*, 102, p. 106301. doi: 10.1016/j.jcomdis.2023.106301.
YORKSTON, K. M., DOWDEN, P. A. ET BEUKELMAN, D. R. (1992): «Intelligibility measurement as a tool in the clinical management of dysarthric speakers», in *Intelligibility in speech Disorders : Theory, measurement and management*. Raymond D. Kent. Madison, Wisconsin: John Benjamins Publishing Company, p. 265-286. Disponible sur: https://benjamins.com/catalog/sspcl.1.08yor.
YORKSTON, K. M., STRAND, E. A. ET KENNEDY, M. R. T. (1996): «Comprehensibility of Dysarthric Speech», *American Journal of Speech-Language Pathology*. American Speech-Language-Hearing Association, 5(1), p. 55-66. doi: 10.1044/1058-0360.0501.55.